



(12) 发明专利

(10) 授权公告号 CN 108984691 B

(45) 授权公告日 2022.01.11

(21) 申请号 201810724536.6

G06F 16/2458 (2019.01)

(22) 申请日 2018.07.04

(56) 对比文件

(65) 同一申请的已公布的文献号
申请公布号 CN 108984691 A

CN 103106280 A, 2013.05.15

CN 106251625 A, 2016.12.21

(43) 申请公布日 2018.12.11

CN 108022006 A, 2018.05.11

CN 106570062 A, 2017.04.19

(73) 专利权人 中国人民解放军陆军工程大学
地址 210008 江苏省南京市秦淮区海福巷1号

CN 106528815 A, 2017.03.22

审查员 周大瑞

(72) 发明人 史涯晴 黄松 郑长友 洪宇
姚奕 惠战伟 王兆丽 董会

(74) 专利代理机构 南京钟山专利代理有限公司
32252

代理人 戴朝荣

(51) Int. Cl.

G06F 16/2453 (2019.01)

权利要求书3页 说明书9页 附图2页

(54) 发明名称

一种基于概率剪枝的路网移动对象范围查询优化方法

(57) 摘要

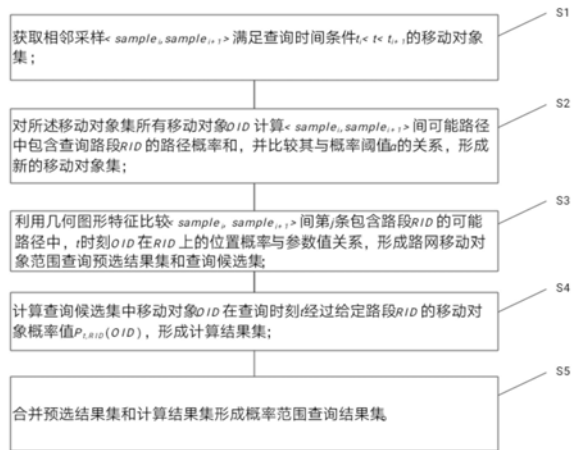
本发明涉及一种基于概率剪枝的路网移动对象范围查询优化方法,属于时空数据技术领域。获取相邻采样<sample_i, sample_{i+1}>满足查询时间条件t_i<t<t_{i+1}的移动对象集;对移动对象集中所有移动对象OID计算<sample_i, sample_{i+1}>间可能路径中包含查询路段RID的路径概率和

$$\sum_{PH} P(v_s, v_e), \text{ 并比较 } \sum_{PH} P(v_s, v_e) \text{ 与概率阈值 } a \text{ 的关系, 形成新的移动对象集; 比较 } < \text{sample}_i, \text{sample}_{i+1}> \text{ 间第 } j \text{ 条包含路段 } RID \text{ 的可能路径中, } t \text{ 时刻 } OID \text{ 在 } RID \text{ 上的位置概率}$$

值a的关系,形成新的移动对象集;比较<sample_i, sample_{i+1}>间第j条包含路段RID的可能路径中,t时刻OID在RID上的位置概率

$$P_{t, RID}^j(OID) \text{ 与 } \frac{\alpha \cdot |PH|}{k} \text{ 关系, 形成路网移动对象范围查询预选结果集和查询候选集; 计算查询候选集中OID在查询时刻 } t \text{ 经过给定路段 } RID \text{ 的移动对象概率值 } P_{t, RID}(O/D), \text{ 形成计算结果集; 合$$

并预选结果集和计算结果集形成概率范围查询结果集。具有可以快速缩小查询候选集、范围查询精度与效率高等优点。



1. 一种基于概率剪枝的路网移动对象范围查询优化方法,其特征在于:包括以下步骤:

(1) 获取相邻采样 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 满足查询时间条件 $t_i < t < t_{i+1}$ 的移动对象集;

(2) 对移动对象集中所有移动对象OID计算 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间可能路径中包含查询路段RID的路径概率和 $\sum_{|PH|} p(v_s, v_e)$, 并比较 $\sum_{|PH|} p(v_s, v_e)$ 与概率阈值 α 的关系,形成新的移动对象集,具体包括:

计算移动对象OID在 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间可能路径中包含查询路段RID的路径概率和 $\sum_{|PH|} p(v_s, v_e)$;

删减移动对象集中 $\sum_{|PH|} p(v_s, v_e) < \alpha$ 的OID,形成新的移动对象集;

其中, $|PH|$ 表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间可能路径的数目, $p(v_s, v_e)$ 表示连续采样 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间某条符合查询时间条件的可能路径中查询路段RID的概率值, v_s 表示查询路段RID的起点, v_e 表示查询路段RID的终点;

(3) 比较 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间第j条包含路段RID的可能路径中,t时刻OID在RID上的位置概率 $P_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 关系,形成路网移动对象范围查询预选结果集和查询候选集;其中,k表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间 $|PH|$ 条可能路径中包含路段RID的路径数目;

(4) 计算查询候选集中移动对象OID在查询时刻t经过给定路段RID的移动对象概率值 $P_{t,RID}(OID)$,形成计算结果集;

(5) 合并预选结果集和计算结果集形成概率范围查询结果集。

2. 根据权利要求1所述的基于概率剪枝的路网移动对象范围查询优化方法,其特征在于:所述获取相邻采样 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 满足查询时间条件 $t_i < t < t_{i+1}$ 的移动对象集具体包括:

如果移动对象OID在时刻t存在确定采样 sample_i 处于查询路段RID上,即 $t = t_i$,则 $P_{t,RID}(OID) = 1$,OID直接形成预选结果集;

如果移动对象相邻 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 的记录时刻 t_i 与 t_{i+1} ,满足 $t_i < t < t_{i+1}$,则 $0 \leq P_{t,RID}(OID) \leq 1$,OID形成移动对象集;此时通过快速判断,将 $P_{t,RID}(OID)$ 一定满足 $P_{t,RID}(OID) \geq \alpha$ 的移动对象OID加入预选结果集,将 $P_{t,RID}(OID)$ 一定满足 $P_{t,RID}(OID) < \alpha$ 的移动对象OID直接剪枝, $P_{t,RID}(OID)$ 值不确定的移动对象OID进一步进行具体概率计算。

3. 根据权利要求1所述的基于概率剪枝的路网移动对象范围查询优化方法,其特征在于:所述移动对象OID在查询时刻t经过给定路段RID的移动对象概率值 $P_{t,RID}(OID)$ 按照下式确定:

$$\begin{aligned}
 P_{t,RID}(OID) &= \sum_{j=1}^k (p(v_s, v_e) \cdot p_{t,RID}^j(OID)) \\
 &= \sum_{j=1}^k \left(\frac{1}{|PH|} \cdot p_{t,RID}^j(OID) \right) ; \\
 &= \frac{1}{|PH|} \cdot \sum_{j=1}^k p_{t,RID}^j(OID)
 \end{aligned}$$

其中, $|PH|$ 表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间可能路径的数目, k 表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间 $|PH|$ 条可能路径中包含路段RID的数目, $p(v_s, v_e)$ 表示连续采样 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间某条符合查询时间条件的可能路径中查询路段RID的概率值, $p_{t,RID}^j(OID)$ 表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间第 j 条包含路段RID的可能路径中 t 时刻OID在RID上的位置概率, $p(v_s, v_e) \cdot p_{t,RID}^j(OID)$ 表示 t 时刻OID在第 j 条满足 $t_m(\text{ph}_j) \leq t_{i+1} - t_i$ 路径上经过路段RID的概率值, $t_m(\text{ph}_j)$ 表示经过 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间某条可能路径 ph_j 的路段最短时间和。

4. 根据权利要求1所述的基于概率剪枝的路网移动对象范围查询优化方法,其特征在于:所述 t 时刻OID在RID上的位置概率 $p_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 关系具体包括:

如果同一移动对象OID的所有可能路径中的 $p_{t,RID}^j(OID)$ 均满足 $p_{t,RID}^j(OID) \geq \frac{\alpha \cdot |PH|}{k}$, 那么 $P_{t,RID}(OID) \geq \alpha$, 则OID必属于查询结果集;

如果同一移动对象OID的所有可能路径中的 $p_{t,RID}^j(OID)$ 均满足 $p_{t,RID}^j(OID) < \frac{\alpha \cdot |PH|}{k}$, 那么 $P_{t,RID}(OID) < \alpha$, 则OID必不属于查询结果集;

除上述以外的情况不能判断 $P_{t,RID}(OID)$ 与 α 的关系, 必须具体计算各个 $p_{t,RID}^j(OID)$ 的精确值。

5. 根据权利要求1所述的基于概率剪枝的路网移动对象范围查询优化方法,其特征在于:所述形成路网移动对象范围查询预选结果集和查询候选集具体包括:

计算 $\frac{\alpha \cdot |PH|}{k}$ 值;

当 $t_{1d}(v_s) \leq t \leq t_{ea}(v_e)$ 时, $p_{t,RID}^j(OID) = 1$;

当 $t_i < t \leq t_{ea}(v_s)$ 或 $t_{1d}(v_e) \leq t < t_{i+1}$ 时, $p_{t,RID}^j(OID) = 0$;

当 $t_{ea}(v_s) < t < t_{1d}(v_s)$ 或 $t_{ea}(v_e) < t < t_{1d}(v_e)$ 时, 判断 $p_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 的关系;

筛选 $t_{ea}(v_s) < t < t_{1d}(v_s)$ 与 $t_{ea}(v_e) < t < t_{1d}(v_e)$ 两种情况中所有 i 值均大于 i_0 , 并且除上述两种情况以外, t 只满足 $t_{1d}(v_s) \leq t \leq t_{ea}(v_e)$ 的OID, 形成预选结果集;

删减 $t_{ea}(v_s) < t < t_{1d}(v_s)$ 与 $t_{ea}(v_e) < t < t_{1d}(v_e)$ 两种情况中所有 i 值均小于 i_0 , 并且除上述两

种情况以外, t 只满足 $t_i < t \leq t_{ea}(v_s)$ 或 $t_{ld}(v_e) \leq t < t_{i+1}$ 的OID, 形成概率范围查询候选集;

其中, 设组成可能路径的各个顶点 v_k 的最早到达时间为 $t_{ea}(v_k)$, 最晚出发时间为 $t_{ld}(v_k)$, $k=0, 1, 2, \dots, n$, 那么 $t_{ea}(v_s)$ 表示查询路段RID起点 v_s 的最早到达时间, $t_{ld}(v_s)$ 表示 v_s 的最晚出发时间, $t_{ea}(v_e)$ 表示查询路段RID终点 v_e 的最早到达时间, $t_{ld}(v_e)$ 表示 v_e 的最晚出发时间, 设 $t_{ea}(v_s) < t < t_{ld}(v_s)$ 时 $i = t - t_{ea}(v_s)$, $t_{ea}(v_e) < t < t_{ld}(v_e)$ 时 $i = t_{ld}(v_e) - t$, i_0 表示最小 i 值。

6. 根据权利要求5所述的基于概率剪枝的路网移动对象范围查询优化方法, 其特征在于: 所述移动对象 t 时刻经过给定路段RID的概率值 $P_{t,RID}^j(OID)$ 包括三种具体情况:

如果 $t_{ld}(v_s) \leq t \leq t_{ea}(v_e)$, 那么OID必定在查询路段RID上, 即 $P_{t,RID}^j(OID) = 1$;

如果 $t_{ea}(v_s) < t < t_{ld}(v_s)$ 或 $t_{ea}(v_e) < t < t_{ld}(v_e)$, 那么OID可能经过查询路段RID, $0 < P_{t,RID}^j(OID) < 1$;

如果 $t_i < t \leq t_{ea}(v_s)$ 或 $t_{ld}(v_e) \leq t < t_{i+1}$, 那么OID必定不在查询路段RID上, 即 $P_{t,RID}^j(OID) = 0$ 。

7. 根据权利要求5所述的基于概率剪枝的路网移动对象范围查询优化方法, 其特征在于: 所述判断 $P_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 的关系具体包括:

计算 i 值, $t_{ea}(v_s) < t < t_{ld}(v_s)$ 时 $i = t - t_{ea}(v_s)$, $t_{ea}(v_e) < t < t_{ld}(v_e)$ 时 $i = t_{ld}(v_e) - t$;

按照 i 值升序排序;

计算 l_{max} ;

计算 i_0 , $i_0 = \frac{\alpha \cdot |PH|}{k} \cdot \frac{l_{max}}{s}$, s 为路段RID的限速;

通过判断移动对象OID所有 i 值与 i_0 之间关系来判断 $P_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 的关系。

8. 根据权利要求7所述的基于概率剪枝的路网移动对象范围查询优化方法, 其特征在于: 所述计算 l_{max} 的具体步骤包括:

计算相邻样本位置 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间某条可能路径 ph_j 的路段最短时间和 $t_m(ph_j)$;

计算 $t_{ea}(v_k)$ 函数与 $t_{ld}(v_k)$ 函数的垂直距离 $t_{ld}(v_k) - t_{ea}(v_k) = t_{i+1} - t_i - t_m(ph_j)$;

查找RID及其所有邻接路段的限速值 s_i ;

取 $(t_{ld}(v_k) - t_{ea}(v_k)) \cdot s_i$ 的最大值作为 l_{max} 。

一种基于概率剪枝的路网移动对象范围查询优化方法

技术领域

[0001] 本发明涉及一种基于概率剪枝的路网移动对象范围查询优化方法,属于时空数据技术领域。

背景技术

[0002] 路网移动对象范围查询是一种典型的移动对象数据库查询类型,在智能交通系统ITS中应用广泛。目前流行的快的打车、滴滴打车等打车软件实质就是进行特定时间特定空间范围内的移动对象范围查询。现有研究大多假设高采样而忽略连续采样间移动对象位置的不确定性,无法满足路网移动对象范围查询的精确度要求。考虑采样频率的移动对象范围查询需要求解每个移动对象所有可能路径中的概率值,涉及大量的概率计算,显然当移动对象数量或者路网密度很大时查询效率较低,远远无法满足查询的实时性要求。因此设计基于概率剪枝的路网移动对象范围查询优化方法,将几何关系与概率判断相结合,兼顾采样频率导致的位置不确定性和数据体量极大性前提下对查询方法进行针对性优化,为基于位置的服务提供精准、实时的路网移动对象范围查询。

[0003] 路网移动对象范围查询的优化方法通常采用空间剪枝策略将空间索引的查询剪枝问题转化为空间距离的快速查询问题。Papadias等人(Papadias D.,Zhang J.,Mamoulis N.and Tao Y.Query processing in spatial network databases[C].Proceedings of the 29th International Conference on Very Large Data Bases,2003:802-813.)将欧氏空间信息和受限网络相结合提出了欧氏限制和网络扩展约束,充分利用位置关联有效缩小搜索空间提高范围查询效率,但此类优化方法未考虑路网移动对象基于采样频率的位置不确定性。Zhang等人(Zhang Y.,Lin X.,Tao Y.,Zhang W.and Wang H.Efficient computation of range aggregates against uncertain location based queries[J].Knowledge and Data Engineering,2012,24(7):1244-1258.)针对查询位置的不确定性,提出了一种基于过滤与判定的方法有效解决了多维空间中的范围聚集查询问题,但该研究的移动对象位置不确定性基于欧氏距离,即两点间的直线距离,相关研究成果并不能直接应用于路网环境。Chen等人(Chen L.,Tang Y.,Lv M.and Chen G.Partition-based range query for uncertain trajectories in road networks[J].Geoinformatica,2015,19(1):61-84.)依据移动对象轨迹单元的网络距离进行划分,提出了一种基于划分的路网移动对象范围查询方法,但这种方法需要进行频繁的磁盘读写操作,大规模路网海量移动对象数据处理的实时性不能满足。

发明内容

[0004] 本发明要解决的技术问题是针对现有路网移动对象范围查询效率不高的缺点,提出一种基于概率剪枝的路网移动对象范围查询优化方法,考虑采样频率导致的路网移动对象位置不确定性的前提下,将几何关系与概率判断相结合,快速缩小查询候选集,提高范围查询的效率。

[0005] 为解决上述技术问题,本发明提供一种基于概率剪枝的路网移动对象范围查询优化方法,包括以下步骤:

[0006] (1) 获取相邻采样 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 满足查询时间条件 $t_i < t < t_{i+1}$ 的移动对象集;

[0007] (2) 对移动对象集中所有移动对象OID计算 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间可能路径中包含查询路段RID的路径概率和 $\sum_{|PH|} p(v_s, v_e)$, 并比较 $\sum_{|PH|} p(v_s, v_e)$ 与概率阈值 α 的关系,形成新的移动对象集。其中, $|PH|$ 表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间可能路径的数目, v_s 表示查询路段RID的起点, v_e 表示查询路段RID的终点;

[0008] (3) 比较 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间第j条包含路段RID的可能路径中,t时刻OID在RID上的位置概率 $P_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 关系,形成路网移动对象范围查询预选结果集和查询

候选集。其中,k表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间 $|PH|$ 条可能路径中包含路段RID的数目;

[0009] (4) 计算查询候选集中移动对象OID在查询时刻t经过给定路段RID的移动对象概率值 $P_{t,RID}(OID)$,形成计算结果集;

[0010] (5) 合并预选结果集和计算结果集形成概率范围查询结果集。

[0011] 进一步地,路网移动对象范围查询定义为:给定以 v_s 为起点、 v_e 为终点的查询路段RID,查询时刻t,概率阈值 α ,路网移动对象范围查询定义为 $q(RID, t, \alpha)$,即返回t时刻所有经过路段RID概率值 $P_{t,RID}(OID) \geq \alpha$ 的移动对象OID集合。

[0012] 进一步地,路网移动对象范围查询结果集存在两种情况:一种是移动对象OID在时刻t存在确定采样 sample_i 处于查询路段RID上,即 $t = t_i$,此时 $P_{t,RID}(OID) = 1$,这种情况OID直接形成预选结果集;

[0013] 另一种是移动对象相邻 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 的记录时刻 t_i 与 t_{i+1} ,满足 $t_i < t < t_{i+1}$,使得t时刻移动对象OID经过路段RID存在一个概率值 $P_{t,RID}(OID)$,满足 $0 \leq P_{t,RID}(OID) \leq 1$;这种情况需要通过快速判断, $P_{t,RID}(OID)$ 一定满足 $P_{t,RID}(OID) \geq \alpha$ 的移动对象加入预选结果集, $P_{t,RID}(OID)$ 一定满足 $P_{t,RID}(OID) < \alpha$ 的移动对象直接剪枝, $P_{t,RID}(OID)$ 值不确定的进一步进行具体概率计算。

[0014] 进一步地,按照下式计算t时刻所有经过路段RID的概率值 $P_{t,RID}(OID)$:

$$\begin{aligned}
 P_{t,RID}(OID) &= \sum_{j=1}^k (p(v_s, v_e) \cdot P_{t,RID}^j(OID)) \\
 [0015] \quad &= \sum_{j=1}^k \left(\frac{1}{|PH|} \cdot P_{t,RID}^j(OID) \right) ; \\
 &= \frac{1}{|PH|} \cdot \sum_{j=1}^k P_{t,RID}^j(OID)
 \end{aligned}$$

[0016] 其中, $|PH|$ 表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间可能路径的数目,k表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间 $|PH|$ 条可能路径中包含路段RID的数目。 $p(v_s, v_e)$ 表示连续采样 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间某条符合查询时间条件的可能路径中查询路段RID的概率值,虽然路段RID可能在多条路径中经过,但是不同路径中路段RID的所处位置或者前后邻接路段不一样,故路段RID的概率

以其所在路径的概率来表示,不能合并。 $P_{t,RID}^j(OID)$ 表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间第j条包含路段RID的可能路径中,t时刻OID在RID上的位置概率。 $p(v_s, v_e) \cdot P_{t,RID}^j(OID)$ 表示t时刻OID在第j条满足 $t_m(\text{ph}_j) \leq t_{i+1} - t_i$ 路径上经过路段RID的概率值,其中 $t_m(\text{ph}_j)$ 表示经过 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间某条可能路径 ph_j 的路段最短时间和。

[0017] 进一步地,比较 $\sum_{|PH|} p(v_s, v_e)$ 与概率阈值 α 的关系形成新的移动对象集具体包括:计算移动对象OID在 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间可能路径中包含查询路段RID的路径概率和 $\sum_{|PH|} p(v_s, v_e)$, $|PH|$ 条可能路径中存在k条包含路段RID的 $\sum_{|PH|} p(v_s, v_e) = \frac{k}{|PH|}$;删减移动对象集中 $\sum_{|PH|} p(v_s, v_e) < \alpha$ 的OID,形成新的移动对象集。

[0018] 进一步地,比较 $P_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 关系具体包括:如果同一移动对象OID的所有可能路径中的 $P_{t,RID}^j(OID)$ 均满足 $P_{t,RID}^j(OID) \geq \frac{\alpha \cdot |PH|}{k}$,那么 $P_{t,RID}(OID) \geq \alpha$,OID必属于查询结果集;如果同一移动对象OID的所有可能路径中的 $P_{t,RID}^j(OID)$ 均满足 $P_{t,RID}^j(OID) < \frac{\alpha \cdot |PH|}{k}$,那么 $P_{t,RID}(OID) < \alpha$,OID必不属于查询结果集;除上面两种以外的情况不能判断 $P_{t,RID}(OID)$ 与 α 的关系,必须具体计算各个 $P_{t,RID}^j(OID)$ 的精确值。

[0019] 进一步地,形成路网移动对象范围查询预选结果集和查询候选集具体包括:计算 $\frac{\alpha \cdot |PH|}{k}$ 值;当 $t_{1d}(v_s) \leq t \leq t_{ea}(v_e)$ 时, $P_{t,RID}^j(OID) = 1$;当 $t_i < t \leq t_{ea}(v_s)$ 或 $t_{1d}(v_e) \leq t < t_{i+1}$ 时, $P_{t,RID}^j(OID) = 0$;当 $t_{ea}(v_s) < t < t_{1d}(v_s)$ 与 $t_{ea}(v_e) < t < t_{1d}(v_e)$ 时,判断 $P_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 的关系;筛选 $t_{ea}(v_s) < t < t_{1d}(v_s)$ ($t_{ea}(v_e) < t < t_{1d}(v_e)$)两种情况中所有i值均大于 i_0 ,并且除上述两种情况以外,t只满足 $t_{1d}(v_s) \leq t \leq t_{ea}(v_e)$ 的OID,形成预选结果集;删减 $t_{ea}(v_s) < t < t_{1d}(v_s)$ ($t_{ea}(v_e) < t < t_{1d}(v_e)$)两种情况中所有i值均小于 i_0 ,并且除上述两种情况以外,t只满足 $t_i < t \leq t_{ea}(v_s)$ 或 $t_{1d}(v_e) \leq t < t_{i+1}$ 的OID,形成概率范围查询候选集。其中, $t_{ea}(v_s)$ 表示查询路段RID起点 v_s 的最早到达时间, $t_{1d}(v_s)$ 表示 v_s 的最晚出发时间, $t_{ea}(v_e)$ 表示查询路段RID终点 v_e 的最早到达时间, $t_{1d}(v_e)$ 表示 v_e 的最晚出发时间,设 $t_{ea}(v_s) < t < t_{1d}(v_s)$ 时 $i = t - t_{ea}(v_s)$, $t_{ea}(v_e) < t < t_{1d}(v_e)$ 时 $i = t_{1d}(v_e) - t$, i_0 表示最小i值。

[0020] 进一步地,当 $t_{ea}(v_s) < t < t_{1d}(v_s)$ 与 $t_{ea}(v_e) < t < t_{1d}(v_e)$ 时,判断 $P_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 的关系具体包括:计算i值, $t_{ea}(v_s) < t < t_{1d}(v_s)$ 时 $i = t - t_{ea}(v_s)$, $t_{ea}(v_e) < t < t_{1d}(v_e)$ 时

$i = t_{ld}(v_e) - t$; 按照 i 值升序排序; 计算 l_{max} ; 计算 $i_0, i_0 = \frac{\alpha \cdot |PH|}{k} \cdot \frac{l_{max}}{s}$, s 为路段RID的限速;

通过判断移动对象OID所有 i 值与 i_0 之间关系来判断 $P_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 的关系。

[0021] 进一步地, 计算 l_{max} 的具体步骤包括: 计算相邻样本位置 $\langle sample_i, sample_{i+1} \rangle$ 间某条可能路径 ph_j 的路段最短时间和 $t_m(ph_j)$; 计算 $t_{ea}(v_k)$ 函数与 $t_{ld}(v_k)$ 函数的垂直距离 $t_{ld}(v_k) - t_{ea}(v_k) = t_{i+1} - t_i - t_m(ph_j)$; 查找RID及其所有邻接路段的限速值 s_i ; 取 $(t_{ld}(v_k) - t_{ea}(v_k)) \cdot s_i$ 的最大值作为 l_{max} 。

[0022] 本发明将几何关系与概率判断相结合, 兼顾采样频率导致的位置不确定性和数据体量极大性前提下对路网移动对象范围查询方法进行针对性优化, 实现了对采样频率导致的轨迹不确定路网移动对象范围查询精度与效率的提高。具有可以快速缩小查询候选集、范围查询精度与效率高等优点。

附图说明

[0023] 图1是本发明的步骤流程图。

[0024] 图2是本发明可能路径顶点的 $t_{ea}(v_k)$ 函数与 $t_{ld}(v_k)$ 函数示意图。

[0025] 图中: $t_i - sample_i$ 对应采样时刻, $t_{i+1} - sample_{i+1}$ 对应采样时刻, v_s - 起点, v_e - 终点, v_k - 顶点, $t_{ea}(v_k)$ - 顶点 v_k 的最早到达时间, $t_{ld}(v_k)$ - 顶点 v_k 的最晚出发时间, $t_{ea}(v_s)$ - 查询路段RID起点 v_s 的最早到达时间, $t_{ld}(v_s)$ - v_s 的最晚出发时间, $t_{ea}(v_e)$ - 查询路段RID终点 v_e 的最早到达时间, $t_{ld}(v_e)$ - v_e 的最晚出发时间, t - 查询时刻 $t, \theta - \tan\theta = l(v_s, v_e) / (t_{ea}(v_e) - t_{ea}(v_s))$, $l(v_s, v_e) - t_{ea}(v_k)$ 函数与 $t_{ld}(v_k)$ 函数中起点 v_s 与终点 v_e 间距离。

具体实施方式

[0026] 下面结合附图对本发明的具体实施方式作进一步详尽描述, 实施例中未注明的技术均为现有技术。

[0027] 实施例1: 参见图1, 本基于概率剪枝的路网移动对象范围查询优化方法包括以下步骤:

[0028] 步骤S1: 获取相邻采样 $\langle sample_i, sample_{i+1} \rangle$ 满足查询时间条件 $t_i < t < t_{i+1}$ 的移动对象集;

[0029] 在本实施方式中, 给定以 v_s 为起点、 v_e 为终点的查询路段RID, 查询时刻 t , 概率阈值 α , 路网移动对象范围查询定义为 $q(RID, t, \alpha)$, 即返回 t 时刻所有经过路段RID概率值 $P_{t,RID}(OID) \geq \alpha$ 的移动对象OID集合。

[0030] 在本实施方式中, 路网移动对象范围查询结果集存在两种情况: 一种是移动对象OID在时刻 t 存在确定采样 $sample_i$ 处于查询路段RID上, 即 $t = t_i$, 此时 $P_{t,RID}(OID) = 1$, 这种情况OID直接形成预选结果集。另一种是移动对象相邻 $\langle sample_i, sample_{i+1} \rangle$ 的记录时刻 t_i 与 t_{i+1} , 满足 $t_i < t < t_{i+1}$, 使得 t 时刻移动对象OID经过路段RID存在一个概率值 $P_{t,RID}(OID)$, 满足 $0 \leq P_{t,RID}(OID) \leq 1$, 这种情况需要通过本实施方式进行快速判断, $P_{t,RID}(OID)$ 一定满足 $P_{t,RID}(OID) \geq \alpha$ 的移动对象加入预选结果集, $P_{t,RID}(OID)$ 一定满足 $P_{t,RID}(OID) < \alpha$ 的移动对象直接

剪枝, $P_{t,RID}(OID)$ 值不确定的进一步进行具体概率计算。

[0031] 具体地, $P_{t,RID}(OID)$ 计算公式如下:

$$\begin{aligned}
 P_{t,RID}(OID) &= \sum_{j=1}^k (p(v_s, v_e) \cdot p_{t,RID}^j(OID)) \\
 [0032] \quad &= \sum_{j=1}^k \left(\frac{1}{|PH|} \cdot p_{t,RID}^j(OID) \right) \quad (1) \\
 &= \frac{1}{|PH|} \cdot \sum_{j=1}^k p_{t,RID}^j(OID)
 \end{aligned}$$

[0033] $|PH|$ 表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间可能路径的数目。 k 表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间 $|PH|$ 条可能路径中包含路段RID的数目。 $p(v_s, v_e)$ 表示连续采样 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间某条符合查询时间条件的可能路径中查询路段RID的概率值。虽然路段RID可能在多条路径中经过,但是不同路径中路段RID的所处位置或者前后邻接路段不一样,故路段RID的概率以其所在路径的概率来表示,不能合并。 $p_{t,RID}^j(OID)$ 表示 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间第 j 条包含路段RID的可能路径中, t 时刻OID在RID上的位置概率。 $p(v_s, v_e) \cdot p_{t,RID}^j(OID)$ 表示 t 时刻OID在第 j 条满足 $t_m(ph_j) \leq t_{i+1} - t_i$ 路径上经过路段RID的概率值,其中 $t_m(ph_j)$ 表示经过 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间某条可能路径 ph_j 的路段最短时间和。

[0034] 步骤S2:对移动对象集中所有移动对象OID计算 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间可能路径中包含查询路段RID的路径概率和 $\sum_{|PH|} p(v_s, v_e)$, 并比较 $\sum_{|PH|} p(v_s, v_e)$ 与概率阈值 α 的关系,形成新的移动对象集;

[0035] 在本实施方式中,步骤S2可以通过以下几个子步骤来完成。

[0036] 步骤S21:计算移动对象OID在 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间可能路径中包含查询路段RID的路径概率和 $\sum_{|PH|} p(v_s, v_e)$, $|PH|$ 条可能路径中存在 k 条包含路段RID的

$$\sum_{|PH|} p(v_s, v_e) = \frac{k}{|PH|};$$

[0037] 步骤S22:删减移动对象集中 $\sum_{|PH|} p(v_s, v_e) < \alpha$ 的OID,形成新的移动对象集;

[0038] 在本实施方式中,因为 $0 \leq p_{t,RID}^j(OID) \leq 1$,由式(1)得 $P_{t,RID}(OID) \leq \sum_{|PH|} p(v_s, v_e)$,

所以只要 $\sum_{|PH|} p(v_s, v_e) < \alpha$ 时,则一定满足 $P_{t,RID}(OID) < \alpha$ 。由步骤S21,对于 $|PH|$ 条可能路径

中存在 k 条包含路段RID的 $\sum_{|PH|} p(v_s, v_e) = \frac{k}{|PH|}$,因此只要求得 $\frac{k}{|PH|} < \alpha$ 就可以直接判断OID

必定不属于路网移动对象范围查询结果集,而不用再计算该移动对象的具体 $p_{t,RID}^j(OID)$

值。

[0039] 步骤S3:比较 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间第j条包含路段RID的可能路径中,t时刻OID在RID上的位置概率 $p_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 关系,形成路网移动对象范围查询预选结果集和查询候选集;

[0040] 在本实施方式中,对于所有步骤2无法处理,即符合条件 $\sum_{|PH|} p(v_s, v_e) \geq \alpha$ 的移动对象,依据查询概率条件 $P_{t,RID}(OID) \geq \alpha$,那么

$$[0041] \quad \frac{1}{|PH|} \cdot \sum_{j=1}^k p_{t,RID}^j(OID) \geq \alpha$$

$$[0042] \quad \sum_{j=1}^k p_{t,RID}^j(OID) \geq \alpha \cdot |PH| \quad (2)$$

[0043] 公式(2)表示OID在k条包含查询路段RID路径上满足查询条件的位置概率值和不小于 $\alpha \cdot |PH|$ 。式子成立的充分条件是 $p_{t,RID}^j(OID) \geq \frac{\alpha \cdot |PH|}{k}$,即OID在任意一条包含查询路段RID路径上满足查询条件的位置概率值不小于 $\frac{\alpha \cdot |PH|}{k}$,因为前提条件是

$1 \geq \sum_{|PH|} p(v_s, v_e) \geq \alpha$, 即 $1 \geq \frac{k}{|PH|} \geq \alpha$, 所以 $1 \geq \frac{\alpha \cdot |PH|}{k} \geq \alpha$, 据此给出基于 $p_{t,RID}^j(OID)$ 的概率剪枝规则:

[0044] (1) 如果同一移动对象OID的所有可能路径中的 $p_{t,RID}^j(OID)$ 均满足 $p_{t,RID}^j(OID) \geq \frac{\alpha \cdot |PH|}{k}$,那么 $P_{t,RID}(OID) \geq \alpha$,OID必属于查询结果集;

[0045] (2) 如果同一移动对象OID的所有可能路径中的 $p_{t,RID}^j(OID)$ 均满足 $p_{t,RID}^j(OID) < \frac{\alpha \cdot |PH|}{k}$,那么 $P_{t,RID}(OID) < \alpha$,OID必不属于查询结果集。

[0046] (3) 除(1)、(2)以外的情况不能判断 $P_{t,RID}(OID)$ 与 α 的关系,必须具体计算各个 $p_{t,RID}^j(OID)$ 的精确值。

[0047] 在本实施方式中,步骤S3可以通过以下几个子步骤来完成。

[0048] 步骤S31:计算 $\frac{\alpha \cdot |PH|}{k}$ 值;

[0049] 步骤S32:当 $t_{ea}(v_s) < t < t_{ld}(v_s)$ 与 $t_{ea}(v_e) < t < t_{ld}(v_e)$ 时,判断 $p_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$

的关系；

[0050] 在本实施方式中,连续采样 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间的某条可能路径 ph_j 上存在若干个路口顶点,表示为 $v_0, v_1, v_2, \dots, v_n$,那么由顶点的最早到达时间和最晚出发时间函数, v_k 的 $t_{ea}(v_k), t_{ld}(v_k)$ 分别计算如下:

$$[0051] \quad t_{ea}(v_k) = t_{ea}(v_{k-1}) + t_m(v_{k-1}, v_k)$$

$$[0052] \quad t_{ld}(v_k) = t_{ld}(v_{k+1}) - t_m(v_k, v_{k+1}) \quad (3)$$

[0053] 其中 $t_{ea}(v_0) = t_i, t_{ld}(v_n) = t_{i+1}$, t_i 表示 sample_i 对应时刻, t_{i+1} 表示 sample_{i+1} 对应时刻。RID为查询路段,起点和终点分别表示为 v_s 和 v_e ,图2表示包含RID的某条可能路径中顶点的最早到达函数 $t_{ea}(v_k)$ 和最晚出发函数 $t_{ld}(v_k)$ 。根据路网移动对象范围查询定义,查询时间限制在 $t_i < t < t_{i+1}$, $P_{t,RID}^j(OID)$ 计算的区域就是 (t_i, t_{i+1}) 范围内。

[0054] 分析图2图形特征,将计算 t 时刻经过给定路段RID的移动对象概率值 $P_{t,RID}^j(OID)$ 分为三种情况:

[0055] (1) 如果 $t_{ld}(v_s) \leq t \leq t_{ea}(v_e)$,那么OID必定在查询路段RID上,即 $P_{t,RID}^j(OID) = 1$;

[0056] (2) 如果 $t_{ea}(v_s) < t < t_{ld}(v_s)$ 或 $t_{ea}(v_e) < t < t_{ld}(v_e)$,那么OID可能经过查询路段RID, $0 < P_{t,RID}^j(OID) < 1$;

[0057] (3) 如果 $t_i < t \leq t_{ea}(v_s)$ 或 $t_{ld}(v_e) \leq t < t_{i+1}$,那么OID必定不在查询路段RID上,即 $P_{t,RID}^j(OID) = 0$ 。

[0058] 在本实施方式中,步骤S32可以通过以下几个子步骤来完成。

[0059] 步骤S321:计算 i 值, $t_{ea}(v_s) < t < t_{ld}(v_s)$ 时 $i = t - t_{ea}(v_s)$, $t_{ea}(v_e) < t < t_{ld}(v_e)$ 时 $i = t_{ld}(v_e) - t$;

[0060] 在本实施方式中,由图2计算 t 时刻的位置概率值 $P_{t,RID}^j(OID) = l_{RID}/l$, l 为 t 直线在 $t_{ea}(v_k)$ 函数与 $t_{ld}(v_k)$ 函数内部分, l_{RID} 为 l 在查询路段RID区域内部分,那么代入

$P_{t,RID}^j(OID) \geq \frac{\alpha \cdot |PH|}{k}$ 进行分析:

$$[0061] \quad P_{t,RID}^j(OID) \geq \frac{\alpha \cdot |PH|}{k}$$

$$[0062] \quad \frac{l_{RID}}{l} \geq \frac{\alpha \cdot |PH|}{k}$$

$$[0063] \quad l_{RID} \geq \frac{\alpha \cdot |PH|}{k} \cdot l \quad (4)$$

[0064] 考虑 l_{RID} 在查询路段RID中计算, $t_{ea}(v_s) < t < t_{ld}(v_s)$ 时 $l_{RID} = (t - t_{ea}(v_s)) \cdot \tan\theta$, $t_{ea}(v_e) < t < t_{ld}(v_e)$ 时 $l_{RID} = (t_{ld}(v_e) - t) \cdot \tan\theta$,由图2中几何图形关系, $\tan\theta = l(v_s, v_e) / (t_{ea}(v_e) - t_{ea}(v_s)) = l(v_s, v_e) / t_m(v_s, v_e) = s$, s 为路段RID的限速。设 $t_{ea}(v_s) < t < t_{ld}(v_s)$ 时 $i = t - t_{ea}(v_s)$, $t_{ea}(v_e) < t < t_{ld}(v_e)$ 时 $i = t_{ld}(v_e) - t$,那么公式(4)就可以表示成:

$$[0065] \quad i \cdot s \geq \frac{\alpha \cdot |PH|}{k} \cdot l$$

$$[0066] \quad i \geq \frac{\alpha \cdot |PH|}{k} \cdot \frac{l}{s} \quad (5)$$

[0067] 步骤S322:按照*i*值升序排序;

[0068] 步骤S323:计算 l_{\max} ;

[0069] 在本实施方式中, l 涉及RID的不同邻接路段,且必须每个移动对象每条可能路径都计算,因此取 l 的统一最大值 l_{\max} , i 只要满足 $i \geq \frac{\alpha \cdot |PH|}{k} \cdot \frac{l_{\max}}{s}$,那么

$$p_{t,RID}^j(OID) \geq \frac{\alpha \cdot |PH|}{k},$$

这样关于 $p_{t,RID}^j(OID)$ 的概率剪枝问题可以转换成基于*i*的时间剪枝问题,具体剪枝规则如下:

[0070] (1) 如果同一移动对象OID的所有可能路径中的*i*均满足 $i \geq \frac{\alpha \cdot |PH|}{k} \cdot \frac{l_{\max}}{s}$,那么

$P_{t,RID}(OID) \geq \alpha$,OID必属于查询结果集;

[0071] (2) 如果同一移动对象OID的所有可能路径中的*i*均满足 $i < \frac{\alpha \cdot |PH|}{k} \cdot \frac{l_{\max}}{s}$,那么

$P_{t,RID}(OID) < \alpha$,OID必不属于查询结果集。

[0072] (3) 除(1)、(2)以外的情况不能判断 $P_{t,RID}(OID)$ 与 α 的关系,必须具体计算各个 $p_{t,RID}^j(OID)$ 的精确值。

[0073] 在本实施方式中,步骤S323可以通过以下几个子步骤来完成。

[0074] 步骤S3231:计算相邻样本位置 $\langle \text{sample}_i, \text{sample}_{i+1} \rangle$ 间某条可能路径 ph_j 的路段最短时间和 $t_m(ph_j)$;

[0075] 步骤S3232:计算 $t_{ea}(v_k)$ 函数与 $t_{ld}(v_k)$ 函数的垂直距离 $t_{ld}(v_k) - t_{ea}(v_k) = t_{i+1} - t_i - t_m(ph_j)$;

[0076] 步骤S3233:查找RID及其所有邻接路段的限速值 s_i ;

[0077] 步骤S3234:取 $(t_{ld}(v_k) - t_{ea}(v_k)) \cdot s_i$ 的最大值作为 l_{\max} 。

[0078] 步骤S324:计算 $i_0, i_0 = \frac{\alpha \cdot |PH|}{k} \cdot \frac{l_{\max}}{s}$, s 为路段RID的限速;

[0079] 步骤S325:判断移动对象OID所有*i*值与 i_0 之间关系,即通过判断移动对象OID所有*i*值与 i_0 之间关系来判断 $p_{t,RID}^j(OID)$ 与 $\frac{\alpha \cdot |PH|}{k}$ 的关系。

[0080] 步骤S33:当 $t_{ld}(v_s) \leq t \leq t_{ea}(v_e)$ 时, $p_{t,RID}^j(OID) = 1$;当 $t_i < t \leq t_{ea}(v_s)$ 或 $t_{ld}(v_e) \leq t < t_{i+1}$ 时, $p_{t,RID}^j(OID) = 0$;

[0081] 步骤S34:筛选 $t_{ea}(v_s) < t < t_{ld}(v_s)$ ($t_{ea}(v_e) < t < t_{ld}(v_e)$)两种情况中所有*i*值均大于

i_0 , 并且除上述两种情况以外, t 只满足 $t_{1d}(v_s) \leq t \leq t_{ea}(v_e)$ 的OID, 形成预选结果集;

[0082] 步骤S35: 删减 $t_{ea}(v_s) < t < t_{1d}(v_s)$ ($t_{ea}(v_e) < t < t_{1d}(v_e)$) 两种情况中所有 i 值均小于 i_0 , 并且除上述两种情况以外, t 只满足 $t_i < t \leq t_{ea}(v_s)$ 或 $t_{1d}(v_e) \leq t < t_{i+1}$ 的OID, 形成概率范围查询候选集。

[0083] 步骤S4: 计算查询候选集中移动对象OID在查询时刻 t 经过给定路段RID的移动对象概率值 $P_{t, RID}(OID)$, 形成计算结果集;

[0084] 步骤S5: 合并预选结果集和计算结果集形成概率范围查询结果集。

[0085] 由上可见, 本发明将几何关系与概率判断相结合, 兼顾采样频率导致的位置不确定性和数据体量极大性前提下对路网移动对象范围查询方法进行针对性优化。本发明实现了对采样频率导致的轨迹不确定路网移动对象范围查询精度与效率的提高。

[0086] 上面结合附图对本发明的技术内容作了说明, 但本发明的保护范围并不限于所述内容, 在本领域的普通技术人员所具备的知识范围内, 还可以在不脱离本发明宗旨的前提下对本发明的技术内容做出各种变化, 凡在本发明的精神和原则之内, 所做的任何修改、等同替换、改进等, 均应包含在本发明的保护范围之内。

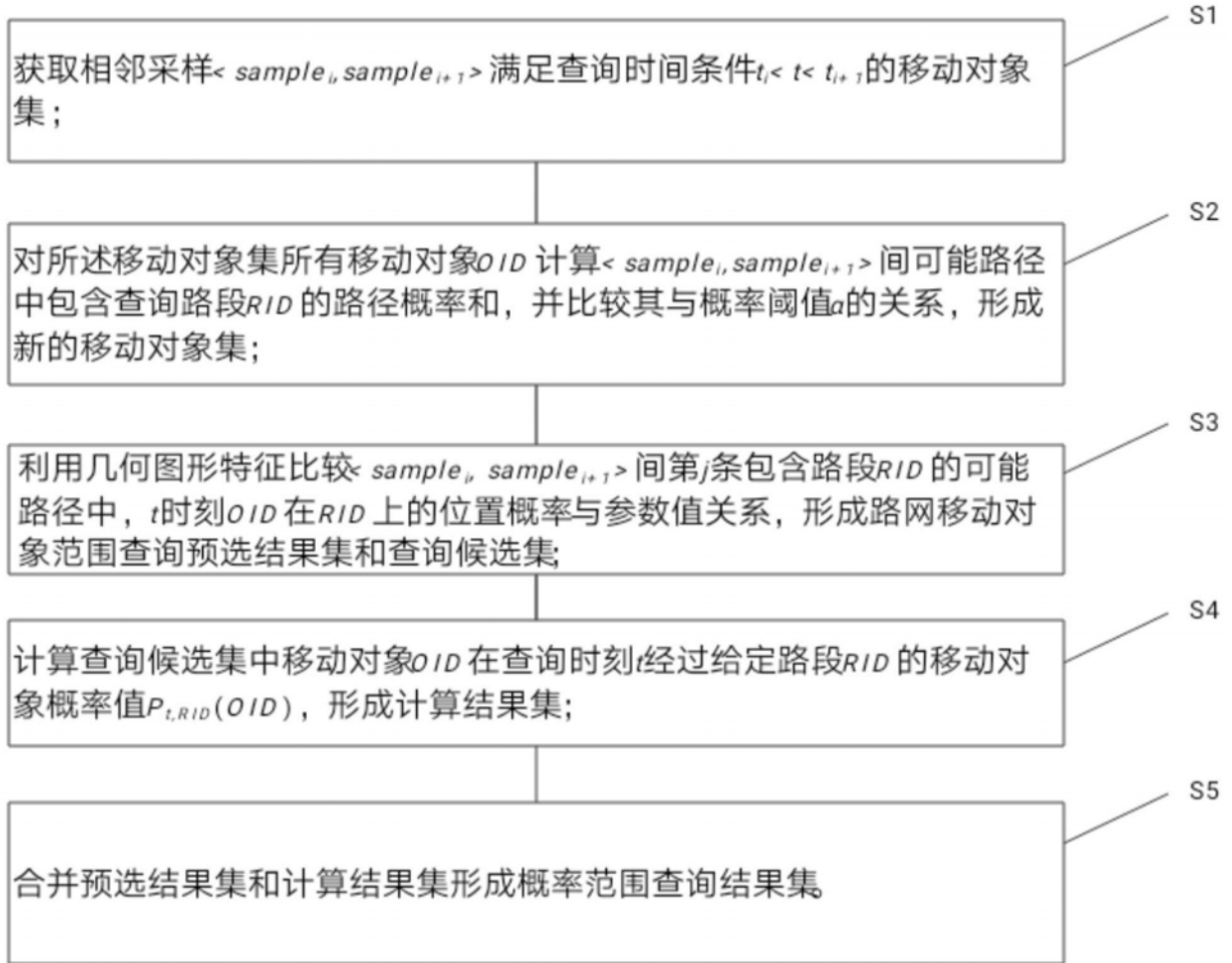


图1

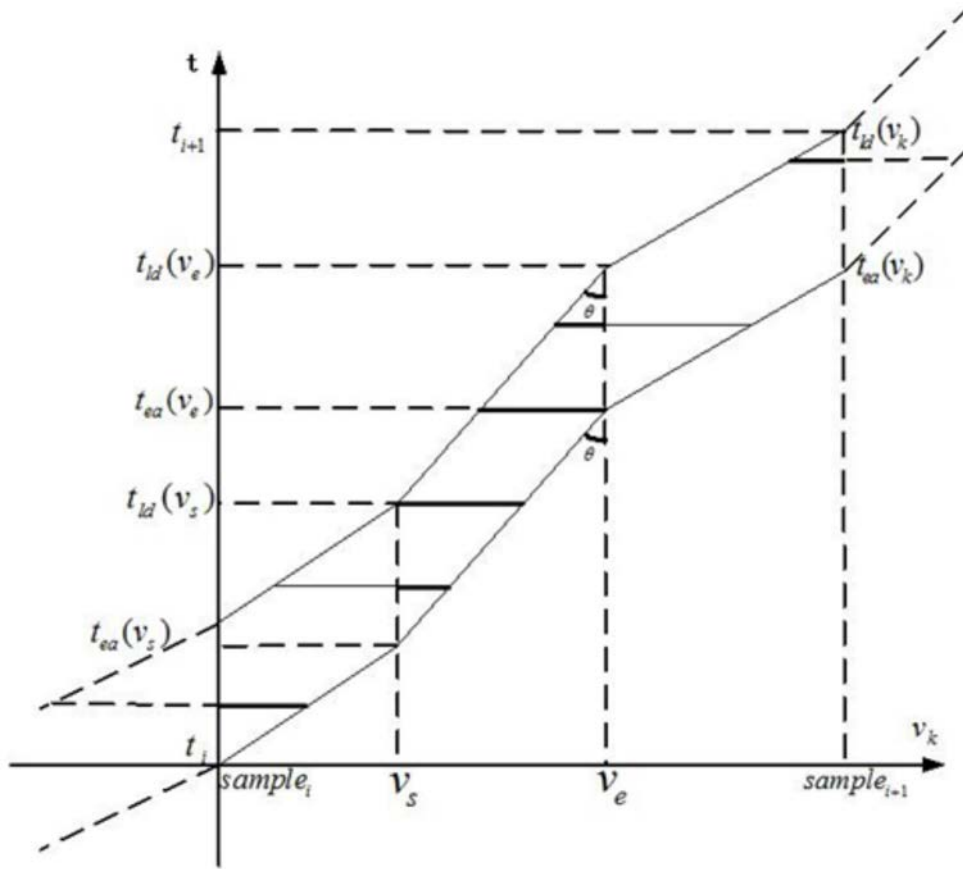


图2